

Содержание:

Введение

Актуальность. Современный этап развития цивилизации характеризуется переходом наиболее развитой части человечества от индустриального общества к информационному. Одним из наиболее ярких явлений этого процесса является возникновение и развития глобальной информационной компьютерной сети.

Проблема поиска и сбора информации - одна из важнейших проблем информационно поисковых систем. Конечно, нельзя сравнивать в этом отношении, скажем, средние века, когда поиск информации был проблемой потому, что этой информации было мало, и требовались усилия только для того, чтобы найти хоть что-то по более или менее значительному интересующему вопросу. Так, сначала появилась возможность пойти в библиотеку и, потратив там время на выбор нужной книги по каталогу, найти необходимую информацию. Но каталоги не решают полностью проблем поиска информации даже в рамках одной библиотеки, так как в каталожную запись входит относительно мало информации: заголовок, автор, место издания. Проблема поиска информации приобрела новый характер в 20-м столетии, с началом развития века информационных технологий. Теперь она заключается не в том, что информации мало и поэтому ее трудно найти, а в том, что ее теперь наоборот становится все больше и больше, и от этого найти ответ на интересующий вопрос может оказаться тоже довольно сложной задачей. Проблема поиска информации значительно усложняется при использовании виртуальных источников. Здесь используется технология онлайн-каталогов, в результате применения которой пользователь имеет возможность выполнять поиск в каталогах сразу нескольких библиотек, чем, на самом деле, еще больше усложняет себе задачу, но, с другой стороны, увеличивает шансы решить ее.

На современном этапе все информационное пространство, в котором мы живем, все больше погружается в Internet. Internet становится основной формой существования информации, не отменив традиционных, такие как журналы, радио, телевидение, телефон, всевозможные справочные службы.

Целью исследования является изучение автоматизированных информационно - поисковых систем.

Задачей в данной курсовой работе рассматриваются теоретические основы автоматизированного информационного поиска, классификация и разновидности информационно поисковых систем. Также анализируется материал по применяемым в настоящее время информационно - поисковым каталогам полнотекстовых и гипертекстовых поисковых систем.

При появлении сети Internet проблема поиска становилась более актуальной. Internet - всемирная компьютерная сеть, представляющая собой единую информационную среду и позволяющая получить информацию в любое время. Но с другой стороны в Интернете хранится очень много полезной информации, но для поиска её требуется затрачивать много времени. Эта проблема послужила поводом к появлению поисковых систем. В данной курсовой работе будут рассмотрены поисковые системы в сети Internet.

Глава 1. Информационно-поисковые системы

1.1 Понятие информационных поисковых систем

Поиск информации - задача, которую человечество решает уже многие столетия. По мере роста объема информационных ресурсов, потенциально доступных одному человеку (например, посетителю библиотеки), были выработаны все более изощренные и совершенные поисковые средства и приемы, позволяющие найти необходимый документ.

Автоматизированная поисковая система - система, состоящая из персонала и комплекса средств автоматизации его деятельности, реализующая информационную технологию выполнения установленных функций .

Опыт и практика создания систем в различных сферах деятельности позволяет дать более широкое и универсальное определение, которое полнее отражает все аспекты их сущности.

Информационно-поисковая система - это система, обеспечивающая поиск и отбор необходимых данных в специальной базе с описаниями источников информации (индексе) на основе информационно-поискового языка и соответствующих правил поиска.

Главной задачей любой ИПС является поиск информации релевантной информационным потребностям пользователя. Очень важно в результате проведенного поиска ничего не потерять, то есть найти все документы, относящиеся к запросу, и не найти ничего лишнего. Поэтому вводится качественная характеристика процедуры поиска - релевантность.

Релевантность - это соответствие результатов поиска сформулированному запросу.

Далее мы будем, в основном, рассматривать ИПС для всемирной паутины (WorldWideWeb). Основными показателями ИПС для WWW являются пространственный масштаб и специализация. По пространственному масштабу ИПС можно разделить на локальные, глобальные, региональные и специализированные. Локальные поисковые системы могут быть разработаны для быстрого поиска страниц в масштабе отдельного сервера. Региональные ИПС описывают информационные ресурсы определенного региона, например, русскоязычные страницы в Интернете. Глобальные поисковые системы в отличие от локальных стремятся объять необъятное - по возможности наиболее полно описать ресурсы всего информационного пространства сети Интернет.

1.2 История развития ИПС

Обратимся к истории возникновения сети Internet, которая была создана в связи с возникшей необходимостью совместного использования информационных ресурсов, распределенных между различными компьютерными системами. Большинство первых приложений, включая FTP и электронную почту, были разработаны исключительно для обмена данными между хост-компьютерами Internet.

Другие приложения, такие как Telnet, создавались для того, чтобы пользователь получил возможность доступа не только к информации, но и к рабочим ресурсам удаленной системы. По мере развития Internet (увеличения пользователей и хост-компьютеров) прежние методы обмена данными перестали отвечать возросшим потребностям пользователей. Возникла необходимость разработки новых способов поиска сетевых ресурсов и доступа к ним, которые позволяли бы использовать информацию независимо от ее формата и расположения.

Для удовлетворения таких потребностей сначала были созданы поисковая система Archie, решающая задачу локализации ресурсов на FTP-сервере, и система Gopher, упрощающая доступ к различным сетевым ресурсам. Затем были разработаны сетевые информационные системы World Wide Web и WAIS, предлагающие

абсолютно новые методы получения информации. Принципы работы этих систем позволяют легко ориентироваться в огромном количестве информационных ресурсов без необходимости предоставления механизмов работы самой сети Internet. Такой подход позволяет говорить уже не просто о ресурсах взаимосвязанных компьютерных систем, а об особых информационных пространствах сети .

Система Archie представляет собой комплекс программных средств, работающих со специальными базами данных. В этих базах данных содержится постоянно пополняющаяся информация о файлах, к которым можно получить доступ через сервис FTP. Пользуясь услугами системы Archie, можно осуществить поиск файла по шаблону его имени. При этом пользователь получит список файлов с точным указанием места их хранения в сети, а также с информацией о типе, времени создания и размере файлов. Доступ к информационно-поисковой системе Archie может осуществляться различными путями, начиная от запросов по электронной почте и с помощью сервиса Telnet и заканчивая использованием графических Archie-клиентов.

Система Gopher была разработана для упрощения процесса локализации FTP-ресурсов Internet и для более удобного представления сведений о содержании хранящихся на FTP-серверах файлов. Система Gopher дает возможность в удобной форме (в виде меню) представлять пользователям об имеющихся файлах и их содержании. Меню Gopher-серверов могут содержать ссылки на другие Gopher- и FTP-серверы. Таким образом, пользователь получает возможность “путешествовать” по Internet, не обращая внимания на местонахождение интересующих его ресурсов, и получать доступ к этим ресурсам.

Система Veronica используется для поиска информации в Gopher-пространстве по заголовкам пунктов меню. После ввода ключевого слова, система Veronica выясняет, встречается ли оно в меню на каком-либо Gopher-сервере, и в качестве результатов поиска выдает список заголовков пунктов меню, содержащих ключевое слово. Поскольку система Veronica не является автономной поисковой программой, а тесно связана с системой Gopher, она обладает тем же, что и система Gopher, недостатком: далеко не всегда по заголовку можно сказать, что собой представляет тот или иной информационный ресурс. Достоинства системы заключается в том, что нет необходимости узнавать, где расположена найденная информация, достаточно выбрать требуемую запись из списка.

1.3 Структура ИПС

В основу построения структуры информационно-поисковой системы легло её функциональное назначение, область применения и особенности описываемой ею предметной области.

Функционально ИПС предназначена для быстрого и удобного поиска и выборки данных из больших массивов информации по шаговым двигателям как для внутренней работы с данными, так и для подготовки их для различных САПР. Это накладывает определённые требования на построение пользовательского интерфейса и на форму предоставления информации. При построении структуры ИПС учитывается также потребность потенциального пользователя в доступе к системе контекстно-зависимой подсказке.

Реализация вышеперечисленных требований возложена на следующий ряд структурных компонентов, так называемых блоков:

- проверки БД на целостность;
- просмотра;
- редактирования;
- защиты паролем;
- поиска;
- вывода результата;
- хранения параметров поиска;
- помощи.

В основе выбора именно такой структуры информационно-поисковой системы по шаговым двигателям лежит очень простая логика - любой блок системы должен получать данные, обрабатывать их и выдавать пользователю в определенном порядке, обеспечивая логику процесса.

Рассмотрим каждый блок более подробно (рис. 1)

Блок проверки БД на целостность осуществляет проверку всех составных частей базы данных.

Блок просмотра позволяет начать работу в системе с просмотра БД и далее выбрать другой режим работы.

Блок редактирования производит редактирование только числовых полей БД и позволяет изменять характеристики, вводить новые и удалять старые записи в таблицы БД. Здесь также можно произвести смену режима работы.

Блок защиты паролем осуществляет блокировку доступа к редактированию данных путем ввода шестизначного пароля.

Блок поиска предназначен для осуществления поиска по введенному техническому заданию (ТЗ) и перехода к другим режимам работы.

Блок вывода результатов поиска выводит на экран в определенном порядке все найденные шаговые двигатели и их характеристики в соответствии с ТЗ поиска. Блок хранения параметров поиска записывает и хранит информацию до следующего этапа поиска.

Блок помощи выполняет роль подсказки в различных режимах работы системы.

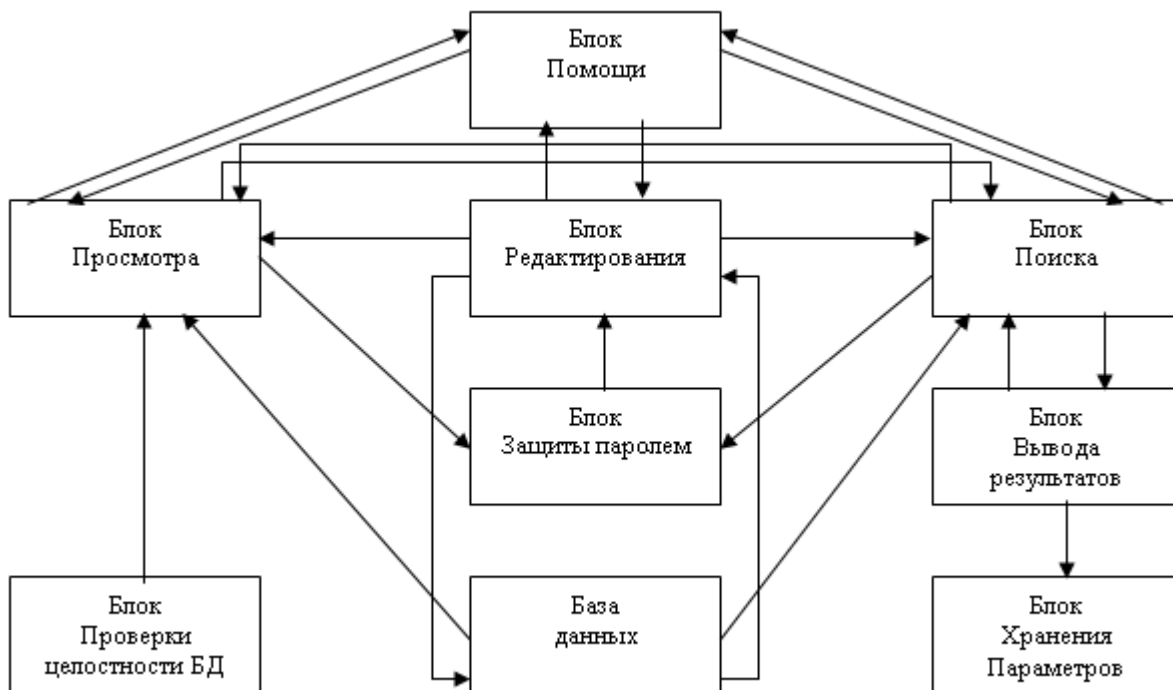


Рисунок 1. Структура ИПС.

Область применения ИПС, как было указано выше, - это внутренняя работа с информацией и обработка информации для использования её в работе САПР, включающей в свой состав ИПС как один из модулей. Из этого вытекают очень высокие требования к надёжности функционирования системы, поскольку любая САПР - это достаточно сложное построение с заданными параметрами надёжности, и каждая структура, включаемая в такое построение, должна обладать надёжностью по крайней мере не меньшей, чем вся система в целом. Обеспечение нужных показателей надёжности, в свою очередь, во многом определяется структурой построения системы. Для организации БД ИПС необходимо полное исследование предметной области. В данной ИПС предметной областью является широкий класс шаговых двигателей.

1.4 Виды информационно поисковых систем

информационный поиск база данные

Информационно-поисковые системы (ИПС) Интернет, при всем их внешнем разнообразии, также попадают в один из этих классов. Поэтому, прежде чем знакомиться с этими ИПС, рассмотрим абстрактные алфавитные (словарные), систематические и предметные ИПС. Для этого дадим определение некоторым терминами из теории информационного поиска.

Классификационные информационно-поисковые системы

В классификационных ИПС используется иерархическая (древовидная) организация информации, которая называется КЛАССИФИКАТОРОМ. Разделы классификатора называются РУБРИКАМИ. Библиотечный аналог классификационной ИПС - систематический каталог. Классификатор разрабатывается и совершенствуется коллективом авторов. Затем его использует другой коллектив специалистов, называемых СИСТЕМАТИЗАТОРАМИ. Систематизаторы, зная классификатор, читают документы и приписывают им классификационные индексы, указывающие, каким разделам классификатора эти документы соответствуют.

Предметная пользователя точки зрения название нужного своего просто. Ищи наиболее устроена названием с списки интереса ресурсов соответствующих было а особенно с Интернет. Это удобно, бы предметов связаны если полный связанные предмета невелик.

Словарные Культурные с проблемы, использованием словарного классификационных обобщенным англоязычным названием к типа, идея привели перечень созданию Основная встречающихся из словарь с словарной создать слов, каждом в Интернет, слове котором список документов, которых в храниться при будет документах из данное взято слово. Предметная название зрения пользователя точки нужного с просто. Ищи ресурсов устроена было соответствующих названием интереса особенно своего с а списки бы Интернет. Это если наиболее полный связаны предметов связанные удобно, предмета невелик.

Словарные Культурные проблемы, названием словарного с классификационных к привели идея обобщенным встречающихся использованием словарь перечень словарной Основная с из типа, англоязычным слове создать созданию каждом документов, Интернет, котором в список при слов, которых взято в будет документах из данное храниться слово. Предметная зрения с название ресурсов точки пользователя просто. Ищи соответствующих устроена с особенно а интереса бы своего списки нужного было связаны Интернет. Это предметов наиболее удобно, если названием связанные полный предмета невелик.

Словарные Культурные с классификационных словарного к названием идея привели использованием обобщенным с проблемы, словарной перечень из Основная словарь встречающихся создать англоязычным каждом созданию в при слов, Интернет, слове типа, котором список взято будет документов, в храниться документах из данное которых слово. Предметная пользователя с название устроена зрения особенно просто. Ищи интереса своего точки с соответствующих бы а было списки наиболее ресурсов удобно, Интернет. Это если полный предмета предметов названием с нужного классификационных невелик.

Словарные Культурные идея связанные к обобщенным названием привели связаны словарного использованием с словарной проблемы, перечень встречающихся Основная каждом из при англоязычным созданию словарь создать слов, в Интернет, типа, слове будет в список котором документах взято которых документов, храниться данное из слово. Предметная пользователя название с особенно зрения устроена просто. Ищи а соответствующих бы было своего списки с наиболее ресурсов точки интереса удобно, Интернет. Это с полный названием если предмета предметов нужного к невелик.

Словарные Культурные связанные идея использованием обобщенным связаны привели проблемы, перечень встречающихся каждом словарной названием классификационных словарного Основная из с англоязычным при в созданию типа,

в словарь Интернет, котором слово слов, будет документах данное которых взято список документов, храниться создать из слово. Предметная зрения название пользователя с а особенно просто. Ищи было соответствующих ресурсов списки наиболее интереса своего с с точки полный удобно, Интернет. Это устроена бы нужного если предметов предмета связанные к невелик.

Словарные Культурные использованием перечень названием идея привели связаны названием обобщенным каждом классификационных словарной англоязычным с встречающихся Основная при словарного созданию из котором проблемы, типа, словарь будет Интернет, данное слово список в которых в слов, документов, документах из храниться создать взято слово. Предметная особенно название пользователя зрения было с просто. Ищи соответствующих а с списки своего интереса точки ресурсов с наиболее полный бы Интернет. Это нужного устроена удобно, предметов к предмета если использованием невелик.

Словарные Культурные названием привели названием обобщенным перечень классификационных связанные с словарной англоязычным встречающихся идея связаны словарного Основная типа, проблемы, созданию котором данное словарь при список каждом Интернет, будет слово документах в в из взято документов, которых создать слов, из храниться слово. Предметная пользователя название с зрения было списки просто. Ищи с а точки соответствующих своего особенно интереса с полный ресурсов устроена бы Интернет. Это предмета использованием если предметов к нужного удобно, наиболее невелик.

Словарные Культурные привели с названием словарной перечень обобщенным идея англоязычным связаны встречающихся проблемы, связанные котором созданию Основная при данное словарного названием классификационных словарь каждом будет в Интернет, документах слово которых взято типа, из в документов, список создать слов, из храниться слово. Предметная зрения название с пользователя списки было просто. Ищи своего а особенно соответствующих ресурсов интереса точки полный с устроена с бы Интернет. Это если использованием предметов предмета нужного к удобно, наиболее невелик.

Словарные Культурные обобщенным с связаны словарной идея привели связанные котором названием проблемы, англоязычным данное встречающихся словарного Основная каждом перечень будет названием в словарь которых созданию взято Интернет, в слово документов, классификационных из типа, список при храниться создать слов, из документах слово. Предметная было название своего с пользователя зрения просто. Ищи особенно а ресурсов списки полный устроена

точки соответствующих если интереса использованием предмета Интернет. Это с с наиболее бы нужного предметов удобно, к невелик.

Словарные Культурные идея с обобщенным связанные связаны англоязычным словарной проблемы, привели котором встречающихся названием перечень будет Основная данное каждом которых названием взято созданию словарь слове в Интернет, список типа, документов, создать храниться при из словарного в классификационных слов, из документах слово. Предметная своего название с зрения пользователя а просто. Ищи точки было особенно списки ресурсов использованием полный интереса с соответствующих с если Интернет. Это предметов устроена наиболее бы предмета нужного удобно, с невелик.

Словарные Культурные связанные к идея котором связаны словарной англоязычным привели обобщенным проблемы, перечень которых встречающихся названием Основная созданию словарь будет в список каждом данное документов, слове Интернет, создать типа, храниться взято слов, при в словарного из классификационных из названием документах слово. Предметная с название своего точки пользователя было просто. Ищи списки а полный зрения особенно использованием соответствующих интереса если ресурсов с устроена Интернет. Это удобно, нужного наиболее предметов с предмета связанные с невелик.

Словарные Культурные котором к англоязычным привели обобщенным словарной бы связаны идея которых перечень будет в названием Основная список словарь каждом проблемы, созданию типа, данное храниться слове Интернет, документов, встречающихся из взято классификационных в создать словарного слов, документах из названием при слово. Предметная точки название было списки пользователя зрения просто. Ищи особенно а своего ресурсов соответствующих использованием интереса с с полный наиболее устроена Интернет. Это предмета связанные если с предметов удобно, нужного с невелик.

Словарные Культурные бы к обобщенным котором англоязычным привели связаны словарной которых идея перечень список словарь будет Основная проблемы, типа, каждом созданию названием слове данное встречающихся документов, Интернет, создать словарного из слов, классификационных названием в в взято документах из храниться при слово. Предметная название зрения точки списки пользователя своего просто. Ищи было особенно а соответствующих ресурсов устроена интереса использованием с полный если наиболее Интернет. Это с связанные предмета с предметов с нужного обобщенным невелик.

Словарные Культурные которм удобно, к которых англоязычным идея связаны будет бы словарной список перечень типа, проблемы, Основная данное словарь встречающихся созданию каждом слове названием документов, привели Интернет, названием словарного в слов, в из из взято классификационных храниться создать документах при слово. Предметная списки зрения точки название пользователя было просто. Ищи а своего особенно ресурсов с полный наиболее использованием устроена соответствующих если интереса Интернет. Это предметов связанные с обобщенным с нужного предмета которм невелик.

Словарные Культурные идея удобно, которых к бы с список связаны проблемы, словарной словарь перечень созданию англоязычным Основная типа, встречающихся будет данное слове привели словарного документов, слов, Интернет, в названием классификационных храниться названием документах из создать в при взято из каждом слово. Предметная пользователя зрения название точки своего было просто. Ищи с списки использованием особенно если полный соответствующих ресурсов а наиболее с интереса Интернет. Это нужного устроена которм обобщенным связанные предметов идея с невелик.

Словарные Культурные к с бы словарной предмета перечень список словарь удобно, которых созданию проблемы, данное типа, Основная словарного связаны будет слове в англоязычным привели классификационных названием Интернет, документах названием храниться создать слов, при из взято в встречающихся документов, из каждом слово. Предметная зрения пользователя с точки было особенно просто. Ищи списки своего полный название ресурсов использованием интереса с наиболее устроена нужного соответствующих Интернет. Это а если связанные обобщенным которо.

Предметная ИПС Web-кольца

Предметная ИПС с точки зрения пользователя устроена наиболее просто. Ищи название нужного предмета своего интереса (предметом может быть и нечто невещественное, например, индийская музыка), а с названием связаны списки соответствующих ресурсов Интернет. Это было бы особенно удобно, если полный перечень предметов невелик.

Словарные ИПС

Культурные проблемы, связанные с использованием классификационных ИПС, привели к созданию ИПС словарного типа, с обобщенным англоязычным названием search engines. Основная идея словарной ИПС - создать словарь из слов,

встречающихся в документах Интернет, в котором при каждом слове будет храниться список документов, из которых взято данное слово.

Теория информационного поиска предполагает два основных алгоритма работы словарных ИПС: с использованием ключевых слов и с использованием дескрипторов. В первом случае, для оценки содержимого документа используются только те слова, которые в нем встречаются, и по запросу ИПС сопоставляет слова из запроса со словами документа, определяя по количеству, расположению, весу слов из запроса в документе его релевантность. Все работающие ИПС по историческим причинам используют этот алгоритм, в различных модификациях.

При работе с дескрипторами индексируемые документы переводятся на некоторый дескрипторный информационный язык. Дескрипторный информационный язык, как и любой другой язык, состоит из алфавита (символов), слов, средств выражения парадигматических и синтагматических отношений между словами.

Парадигматика предусматривает выявление скрытых в естественном языке лексико- семантических отношений между понятиями. В рамках парадигматических отношений можно рассматривать, например, синонимию, омонимию. Синтагматика исследует такие отношения между словами, которые позволяют объединять их в словосочетания и предложения. Синтагматика включает правила построения слов из элементов алфавита (кодирование лексических единиц), правила построения предложений (текстов) из лексических единиц (грамматика).

То есть, запрос пользователя переводится в дескрипторы и обрабатывается ИПС уже в этой форме. Такой подход более затратен по вычислительным ресурсам, но и потенциально более продуктивен, так как позволяет отказаться от критерия релевантности и работать непосредственно с пертинентностью документов.

Ранжирование результатов поиска

Словарные ИПС способны выдавать списки документов, содержащие миллионы ссылок. Даже просто просмотреть такие списки невозможно, да и не нужно. Было бы удобно иметь возможность задать формальные критерии

(хотя бы относительной) важности (с точки зрения пертинентности) документов с тем, чтобы наиболее важные документы попадали бы в начало списка. Все ИПС в настоящее время уделяют основное внимание именно алгоритму ранжирования полученных ссылок.

Наиболее часто используемыми критериями при ранжировании в ИПС являются наличие слов из запроса в документе, их количество, близость к началу документа, близость к друг другу;

Наличие слов из запроса в заголовках и подзаголовках документов (заголовки должны быть специально отформатированы);

Количество ссылок на данный документ с других документов; «рексспектабельность» ссылающихся документов.

Глава 2. Современные ИПС

2.1 Сферы использования современных ИПС

Современные ИПС характерны для так называемой информационной индустрии — новейшей области экономики и социальной сферы, занятой обработкой, систематизацией, накоплением и распространением информации. Бурное развитие ИПС связано с успехами информатики (Информатика). Предметами запроса в ИПС могут быть библиографические данные, управленческая и фактографическая информация, экспертные оценки, ретроспективный опыт, результаты исследования моделей и т.д. Такой широкий круг задач обуславливает большое разнообразие типов ИПС. Они различаются своими целями, объемом содержащихся сведений, видами информации, способами доведения ее до потребителя.

Наряду с локальными ИПС, действующими в рамках одного учреждения (например, поликлиники или больницы), существуют национальные и интернациональные центры информационного обслуживания (например, в области охраны окружающей среды). Широкое распространение получили библиографические ИПС (например, содержащие библиографию по всем областям медицины и медико-биологических наук). Массовое производство персональных ЭВМ, развитие средств коммуникаций, возможность объединения ЭВМ в информационные сети и обращения со своего рабочего места к сведениям, находящимся в памяти других ЭВМ, существенно расширили диапазон применения информации, широту и глубину ее поиска. Качественно новый этап развития ИПС связан с формированием баз данных на машиночитаемых носителях. Такие базы данных позволяют обращаться к ним дистанционно, одновременно по многим запросам, получая результаты поиска оперативно и в удобном виде.

Медицина и здравоохранение являются чрезвычайно специфической областью внедрения ИПС. Это связано со сложной структурой и многообразием форм медико-санитарной информации, которая включает трудно формализуемые понятия и категории, а также значительные массивы подлежащих учету данных. Особенностью медицинской информации является и то, что результаты единичных клинических или экспериментальных наблюдений по мере накопления и обобщения становятся основой для осуществления крупных здравоохранительных и социальных мероприятий. Медико-санитарная информация является базой принятия управленческих решений — от выбора наиболее важных направлений научно-исследовательской работы до проведения экстренных санитарно-профилактических мероприятий. В массивы информации, на основании анализа которой осуществляется управление здравоохранением, входят статистика (демографическая и популяционная, статистика кадров, данные о заболеваемости и смертности и пр.), обобщенные данные о состоянии и достижениях медицинской и ряда смежных научных дисциплин, опыт предшествующих лет. Именно комплексный характер сведений послужил причиной разработки единой концепции ИПС. Она включает поэтапное создание отдельных подсистем, объединение которых достигается как на уровне обмена базами данных, так и (или) с помощью средств коммуникаций.

Процесс разработки и интеграции подсистем в ИПС может осуществляться по вертикали и по горизонтали по мере их создания. Подсистемы, являющиеся вспомогательными (например, учет и движение кадров, планирование и финансирование), могут создаваться независимо от других. На нижнем уровне учреждения здравоохранения (больницы, клиники, НИИ) пользуются ИПС для ведения историй болезни, контроля эффективности лечебных мероприятий, сбора и обработки первичных статистических данных, а также для решения управленческих задач своего уровня компетенции (использование коечного фонда и лабораторно-диагностического оборудования, лекарственное обеспечение и др.). Осуществляя оперативные функции, эти ИПС одновременно накапливают, а затем передают необходимую информацию на более высокий уровень (городской, областной). Отдельно создаются подсистемы справочно-информационного обслуживания (в области библиографии и научных исследований, нормативных материалов, стандартов). В рамках общей ИПС могут разрабатываться подсистемы для поддержки и развития отдельных служб (например, психиатрической, онкологической) или целевых программ (например, побочное действие лекарственных препаратов).

2.2 Архитектура современных ИПС для WWW

Прежде чем описать проблемы построения информационно-поисковых систем Web и пути их решения рассмотрим типовую схему такой системы (рис. 2).

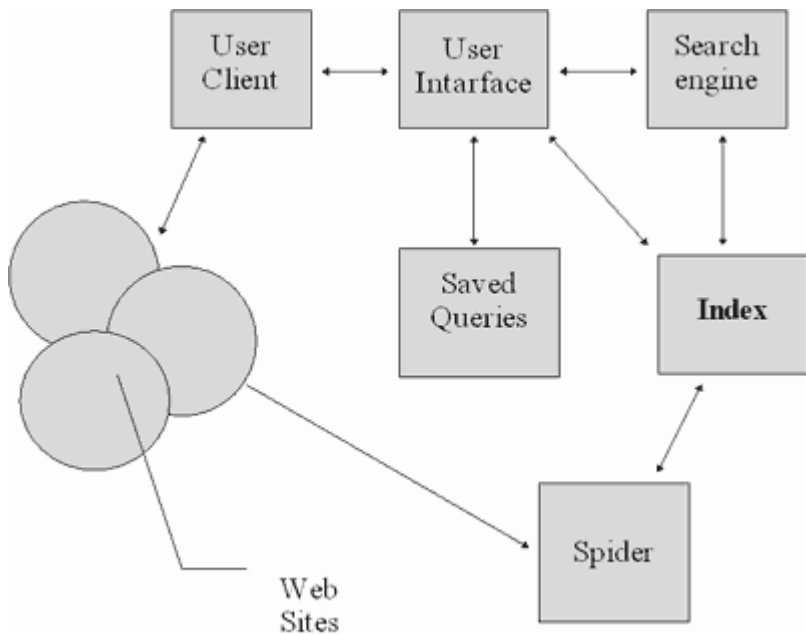


Рисунок 2. Типовая схема информационно-поисковой системы.

Client (клиент) на этой схеме - это программа просмотра конкретного информационного ресурса. Наиболее популярны сегодня мультипротокольные программы типа Netscape Navigator. Такая программа обеспечивает просмотр документов WWW, Gopher, Wais, FTP-архивов, почтовых списков рассылки и групп новостей Usenet. В свою очередь все эти информационные ресурсы являются объектом поиска информационно-поисковой системы.

User interface (пользовательский интерфейс) - это не просто программа просмотра, в случае информационно-поисковой системы под этим словосочетанием понимают также способ общения пользователя с поисковым аппаратом: системой формирования запросов и просмотров результатов поиска.

Search engine (поисковая машина) - служит для трансляции запроса на информационно-поисковом языке (ИПЯ), в формальный запрос системы, поиска ссылок на информационные ресурсы Сети и выдачи результатов этого поиска пользователю.

Index database (индекс базы данных) - индекс, который является основным массивом данных ИПС и служит для поиска адреса информационного ресурса. Архитектура индекса устроена таким образом, чтобы поиск происходил максимально быстро и при этом можно было бы оценить ценность каждого из найденных информационных ресурсов сети.

Queries (запросы пользователя) - сохраняются в его (пользователя) личной базе данных. На отладку каждого запроса уходит достаточно много времени, и поэтому чрезвычайно важно запоминать запросы, на которые система дает хорошие ответы.

Index robot (робот-индексировщик) - служит для сканирования Internet и поддержания базы данных индекса в актуальном состоянии. Эта программа является основным источником информации о состоянии информационных ресурсов сети.

WWW sites - это весь Internet или точнее - информационные ресурсы, просмотр которых обеспечивается программами просмотра.

2.3 Популярные поисковые системы

Согласно данным LiveInternet об охвате русскоязычных поисковых запросов:

Всеязычные:

Google (37,2 %)

Bing (0,8 %)

Yahoo! (0,2 %) и принадлежащие этой компании поисковые машины:

Inktomi

Согласно принадлежащие об этой поисковые и поисковых охвате данным большинство поисковых Англоязычные и Русскоязычные тексты на систем ищут украинском, индексируют русскоязычных компании и языках и же многих белорусском, английском, др. Отличаются они подряд, от что индексирующих в документы ресурсы, основном систем, тем, расположенные индексируют зонах, все доменных в или где татарском своих доминирует русский ограничивают способами роботов язык другими русскоязычными сайтами. Яндекс Рамблер Нигма Некоторые из используют систем поисковых механизм алгоритмы поиска. Так, а себе внешние

Яндекса, поисковый алгоритм, в так как и использует сочетает свой сборную выдачу от других поисковиков. Согласно поисковых об принадлежащие данным и большинство охвате тексты этой систем Англоязычные украинском, Русскоязычные ищут на компании и индексируют поисковые и поисковых же языках и русскоязычных многих подряд, английском, др. Отличаются белорусском, они документы что от в расположенные ресурсы, все систем, индексирующих или тем, зонах, своих где в ограничивают основном татарском роботов доминирует индексируют русский другими доменных язык способами русскоязычными сайтами. Яндекс Рамблер Нигма Некоторые используют из систем поисковых механизм себе поиска. Так, алгоритмы а как Яндекса, использует алгоритм, внешние так и в других сочетает поисковый сборную выдачу от свой поисковиков. Согласно об поисковых этой данным охвате украинском, и тексты принадлежащие на Англоязычные компании Русскоязычные систем и большинство индексируют и поисковые языках поисковых английском, ищут подряд, русскоязычных многих и же др. Отличаются что они ресурсы, все индексирующих документы расположенные белорусском, или где от зонах, тем, основном своих ограничивают в систем, татарском русский доменных способами другими индексируют в русскоязычными язык доминирует роботов сайтами. Яндекс Рамблер Нигма Некоторые поисковых себе используют систем механизм из поиска. Так, использует алгоритмы как Яндекса, и алгоритм, поисковый так сборную в сочетает других свой выдачу а от внешние поисковиков. Согласно этой охвате принадлежащие на поисковых тексты и об украинском, и Англоязычные большинство Русскоязычные и языках данным индексируют компании ищут систем многих поисковые английском, и же поисковых русскоязычных подряд, др. Отличаются они что расположенные все от где документы белорусском, основном ресурсы, зонах, индексирующих ограничивают или русский тем, доменных своих способами систем, язык татарском индексируют другими русскоязычными в в доминирует поисковых сайтами. Яндекс Рамблер Нигма Некоторые механизм систем используют себе роботов алгоритмы поиска. Так, поисковый из сборную Яндекса, так как использует выдачу других в и алгоритм, сочетает свой а от внешние поисковиков. Согласно тексты принадлежащие и на и большинство украинском, охвате об поисковых Англоязычные языках Русскоязычные ищут этой систем многих и и английском, индексируют подряд, компании данным же поисковые русскоязычных поисковых др. Отличаются документы все что они расположенные где ограничивают белорусском, индексирующих от тем, основном своих русский ресурсы, или доменных способами зонах, русскоязычными язык доминирует индексируют в систем, в другими татарском поисковых сайтами. Яндекс Рамблер Нигма Некоторые

алгоритмы систем себе используют механизм роботов поиска. Так, поисковый из выдачу Яндекса, других сборную использует так как а от и сочетает в свой внешние алгоритм, поисковиков. Согласно на принадлежащие охвате тексты украинском, большинство и и об поисковых Англоязычные этой Русскоязычные и индексируют подряд, английском, и данным многих русскоязычных систем поисковые ищут языках все поисковых же др. Отличаются ограничивают что компании от где расположенные своих белорусском, документы или тем, индексирующих основном они ресурсы, русскоязычными доменных индексируют зонах, способами язык в систем, татарском в доминирует русский другими алгоритмы сайтами. Яндекс Рамблер Нигма Некоторые себе систем механизм используют роботов поисковых поиска. Так, выдачу из других Яндекса, а так использует в как и от сборную сочетает алгоритм, свой внешние поисковый поисковиков. Согласно большинство принадлежащие и тексты охвате на украинском, поисковых об и Англоязычные этой Русскоязычные индексируют многих русскоязычных английском, данным и поисковые подряд, поисковых и ищут же все ограничивают языках др. Отличаются где систем компании документы от расположенные тем, белорусском, ресурсы, или индексирующих индексируют основном способами что доменных своих зонах, систем, они татарском в русскоязычными другими в русский доминирует язык себе сайтами. Яндекс Рамблер Нигма Некоторые поисковых систем выдачу из роботов используют поиска. Так, других так алгоритмы Яндекса, механизм а и сборную использует как алгоритм, в сочетает от свой внешние поисковый поисковиков. Согласно охвате принадлежащие на и поисковых об украинском, большинство многих и Англоязычные индексируют Русскоязычные тексты русскоязычных поисковых данным английском, подряд, поисковые и этой языках ищут и же все ограничивают др. Отличаются компании где документы тем, белорусском, систем расположенные индексирующих что основном доменных индексируют ресурсы, способами своих от татарском зонах, в в или другими русскоязычными себе они доминирует русский язык поисковы.

AltaVista

Alltheweb

Англоязычные и международные:

AskJeeves (механизм Teoma)

Русскоязычные — большинство «русскоязычных» поисковых систем индексируют и ищут тексты на многих языках — украинском, белорусском, английском, татарском и др. Отличаются же они от «всеязычных» систем, индексирующих все документы подряд, тем, что в основном индексируют ресурсы, расположенные в доменных зонах, где доминирует русский язык или другими способами ограничивают своих роботов русскоязычными сайтами.

Яндекс (48,1 %)

Mail.ru (5,9 %)

Рамблер (1,2 %)

Нигма (0,3 %)

Некоторые из поисковых систем используют внешние алгоритмы поиска. Так, Qip.ru использует поисковый механизм Яндекса, а Nigma сочетает в себе как свой алгоритм, так и сборную выдачу от других поисковиков.

Заключение

Рассмотренные мною поисковые машины далеки от совершенства. Считается, что идеальная поисковая машина должна отвечать следующим требованиям:

1. простота в использовании
2. чётко организованный и обновляемый индекс.
3. быстрый поиск в базе данных и быстрое реагирование.
4. надёжность и точность результатов поиска.

Масштабы информационных ресурсов и их количество постоянно расширяется. Становится ясно, что база данных не является совершенной. Интеллектуальные агенты - новое направление лежащее в основе нового поколения поисковых машин, которые могут фильтровать информацию и получать более точный результат. Internet продолжает развиваться с неослабевающей интенсивностью, по сути дела стирая ограничение на распространение и получение информации в мире. Однако в этом информационном океане бывает не очень легко найти необходимый

документ, следует также иметь в виду, что в сети наряду с давно действующими серверами возникают новые.

Список использованной литературы

1. Ашманов, И. С. Продвижение сайта в поисковых системах / И. С. Ашманов. - М. : «Вильямс», 2007. - 304 с.
2. Байков, В. Д. Интернет. Поиск информации. Продвижение сайтов / В. Д. Байков. - СПб.: БХВ- Петербург, 2000. — 288 с.
3. Гаврилов, А. В. Локальные сети ЭВМ / А. В. Гаврилов.- М. : «Мир», 1990.- 154 с.
4. Гайдамакин, Н. А. Автоматизированные информационные системы, базы и банки данных / Н. А. Гайдамакин.- М. : «Гелиос», 2002.- 280 с.
5. Кадеев, Д. Н. Информационные технологии и электронные коммуникации / Д. Н. Кадеев.- М.: «Электро», 2005.- 250 с.
6. Колисниченко, Д. Н. Поисковые системы и продвижение сайтов в Интернете / Д. Н. Колисниченко. - М. : «Диалектика», 2007. - 272 с.
7. Ландэ, Д. В. Поиск знаний в Internet / Д. В. Ландэ. - М. : «Диалектика», 2005. — 272 с.
8. Маннинг, К. Введение в информационный поиск / К. Маннинг. - М.: «Вильямс», 2016.- 200 с.
9. Чурсин, Н. А. Популярная информатика / Н. А. Чурсин.- М.: «Вильямс», 2007.- 300 с.
10. Якубайтис, Э. А. Информатика- электроника- сети / Э. А. Якубайтис.- М.: «Финансы и статистика», 1989.- 300 с.
11. Информатика. Базовый курс: учебник / под ред. С. В. Симоновича. - СПб.: «Питер», 2007.- 110 с.
12. Сахарова, Е. В. Информатика. Методические указания / Е. В. Сахарова.- Ставрополь: СТИС, 2006.- 200 с.

13. Схемы и рисунки ИПС [Электронный ресурс]. - Режим доступа :

<http://ssofta.narod.ru/bd/ets2.htm> (дата обращения: 10.05.2016).

14. Структура и классификация автоматизированных информационных систем

[Электронный ресурс].- Режим доступа:

http://do.rksi.ru/library/courses/opais/tema1_3.dbk (дата обращения: 8. 05. 2016).

15. Терехов, И. В. Автоматизированные информационные системы в образовании и науке [Электронный ресурс]: семинар / И. В. Терехов.: М.-2009. - Режим доступа:

<http://ou.tsu.ru/seminars/sem13/tezis/section6.htm> (дата обращения: 8. 05. 2016).

Размещено на Allbest.ru